Technical Appendix

We used a bootstrap procedure [1] to estimate the variability in panel members' scores, and hence a proposal's rank, and hence the variability in funding decisions. The steps of the bootstrap were:

1. For each proposal the panel members' scores were randomly re-sampled from the original scores with replacement. This can be thought of as the same proposal being re-reviewed by a slightly different panel (e.g., in a different year)
2. The re-sampling was repeated 1,000 times for every proposal to create 1,000 alternative realities
3. For each alternative reality the proposals were ranked in each panel, and the same number of proposals were awarded per panel as for the original decision

We illustrate the bootstrap procedure using a small hypothetical example below.

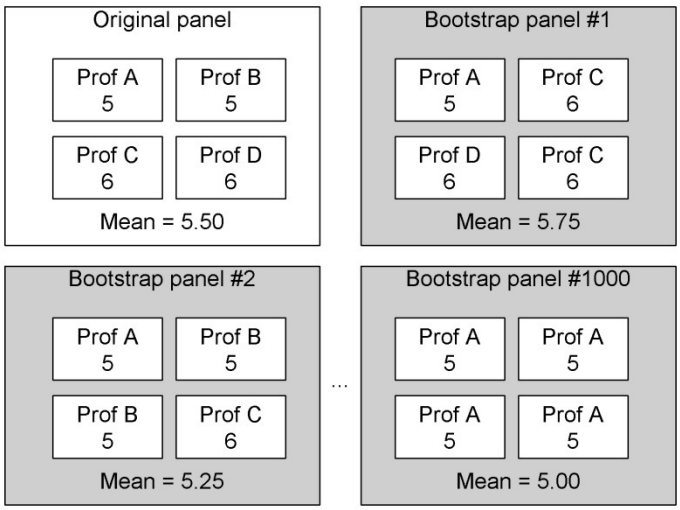

Figure: Hypothetical example of the bootstrap procedure for one proposal using a panel of four members

The original panel gave the grant a mean score of 5.50, based on two 5's and two 6's. In the first bootstrap panel Professor C was randomly selected twice, whereas Professor B was not selected. This meant the proposal did better than its original score with a mean score of 5.75. In the 1,000th bootstrapped panel Professor A was selected each time, meaning the proposal did worse than its original score with a mean of 5.00.

The bootstrap procedure estimates the variance in observed scores. This bootstrap procedure is non-parametric, which is essential here because we could not safely assume a Normal distribution for

the panel scores. Similar simulation methods have been used to assess the reliability of ranking hospitals according to their quality of the care [2].

1.Davison, A.C. and D.V. Hinkley, Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics. 1997, Cambridge: Cambridge University Press.

2. Marshall, E.C., et al., Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates / Commentary: How robust are rankings? The implications of confidence intervals. BMJ, 1998. 316(7146): p. 1701-1705.